

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

Vision Research 46 (2006) 2901–2913

**Vision  
Research**
[www.elsevier.com/locate/visres](http://www.elsevier.com/locate/visres)

# Can the theory of “whitening” explain the center-surround properties of retinal ganglion cell receptive fields?

Daniel J. Graham \*, Damon M. Chandler, David J. Field

Department of Psychology, Cornell University, Ithaca, NY 14853, USA

Received 9 February 2005; received in revised form 6 March 2006

## Abstract

To account for the spatial and temporal response properties of the retina, a number of studies have proposed that these properties serve to “whiten” the visual input. In particular, it has been argued that the sensitivity of retinal ganglion cells is matched to the spatial frequency spectrum of natural scenes, resulting in a flattened or “whitened” response spectrum across a range of frequencies. However, we argue that there are two distinct hypotheses regarding the flattening of the spectrum. The *decorrelation hypothesis* proposes that the magnitude of each ganglion cell tuning curve rises with spatial frequency, resulting in a flattened response spectrum for natural scene stimuli. With appropriate sampling, this scheme allows neighboring neurons to be uncorrelated with each other. The *response equalization hypothesis* proposes that the overall response magnitude of neurons increases with spatial frequency. The proposed goal of this model is to allow neurons with different receptive field sizes to produce the same average response to natural scenes. The response equalization hypothesis proposes an explanation for the relative gain of different ganglion cells and we show that this proposal fits well with published data. We suggest that both hypotheses are important in understanding the tuning and sensitivity of ganglion cells. However, using a simulation, both models are shown to be insufficient to explain the center-surround receptive field organization of ganglion cells. We discuss other factors, including representational sparseness, which could be related to the goals of ganglion cell spatial processing. We suggest three constraints needed to describe the basic linear properties of P-type ganglion cells: decorrelation, response equalization, and a minimal wiring or minimal size constraint.

© 2006 Elsevier Ltd. All rights reserved.

**Keywords:** Retina; Vision; Natural scenes; Retinal coding; Center-surround; Receptive field; Ganglion cell; Spatial vision; Vector length; Contrast; Response equalization; Sparse coding; Nonlinear; Minimum wiring

## 1. Introduction

Across mammalian species one finds that the layout, development, and structure of the retina is remarkably well-conserved (Finlay, de Lima Silveira, & Reichenbach, 2004). Center-surround antagonism in particular is found in some form in the early visual systems of all vertebrates and invertebrates (e.g., Land, 1985). Although there have been numerous proposals regarding what is achieved by retinal processing—ranging from “edge enhancement” (Balboa & Grzywacz, 2000; Ratliff, 1965) to decorrelation—we will argue that current models of retinal ganglion

cells are insufficient to account for basic aspects of information processing in the retina. We also wish to emphasize that the account we provide here is in no way complete. The varieties of non-linearities and cell types will certainly require a much richer model. Our emphasis, however, is on the basic center-surround organization of retinal ganglion cells.

### 1.1. “Whitening” and decorrelation

Natural scenes typically have strong spatial pairwise correlations which can be expressed as a spatial frequency amplitude spectrum that falls as  $1/\text{frequency}$ —or as a power spectrum that falls as  $1/f^2$  (Burton & Moorhead, 1987; Field, 1987). The prevailing view of retinal coding

\* Corresponding author. Fax: +1 607 255 8433.

E-mail address: [djg45@cornell.edu](mailto:djg45@cornell.edu) (D.J. Graham).

was shaped by Srinivasan, Laughlin, and Dubs (1982), and by Atick and Redlich (1992), whose *decorrelation hypothesis* focuses on the relation between the ganglion cell tuning curves and the spectra of natural scenes.

Srinivasan et al. (1982) showed that a group of detectors sampling over space will transmit the greatest amount of information given the presence of noise by taking a weighted linear sum over the spatial arrangement of a group of detectors. Atick and Redlich (1992) extended this line of thinking and proposed that the goal of retinal coding is to produce a decorrelated output in response to natural scenes. They noted that for a range of spatial frequencies, the tuning curve for ganglion cells increases with spatial frequency. Since natural scenes' amplitude spectra fall with increasing spatial frequency, the multiplication of these two spectra (which corresponds to a spatial convolution) will result in a flat spectrum over this range of frequencies. Flattening of the spectrum is sometimes called "whitening" and it can result in neurons with decorrelated activity under appropriate sampling conditions.

However, two quite separate ideas of whitening have been proposed. Both address the ways in which the early visual system handles the  $1/f$  amplitude spectrum of natural scenes but each has different requirements and each achieves different objectives:

- The first theory of whitening, which we will call the *response equalization hypothesis*, was proposed by Field (1987) for cortical neurons, and extended by Brady and Field (1995, 2000) and Field and Brady (1997). In this account, the goal is to produce a representation where each neuron has roughly the same average activity in the presence of natural scenes. Neurons tuned to high frequencies would need increased response gain to produce the same response as low frequency neurons.
- As described above, the *decorrelation hypothesis* of Atick and Redlich (1992) argues that the relationship between the spectrum of each individual ganglion cell and the  $1/f$  spectrum of the input results in decorrelated responses. This decorrelation depends on both the relative spectra and the sampling density of neurons.

These two models are not incompatible with each other (see Appendix A). Indeed, both can be independently correct or incorrect. The decorrelation hypothesis is appealing because it predicts spatial redundancy reduction at the retinal output. With appropriate retinal sampling, neighboring neurons will have no pairwise correlations in space. In the case of retinal neurons, the center-surround structure of the filters results in bandpass tuning curves for which a portion of the curve rises with frequency. As Atick and Redlich note, the increase in sensitivity with frequency has one important cost: it magnifies the noise at high frequencies. They provide a convincing argument that the reduction in sensi-

tivity at the higher spatial frequencies, especially under low light conditions (i.e., high noise) provides an efficient strategy for coding natural scenes. This line of argument requires two important features: (1) the tuning curves must have the appropriate shape, and (2) the tiling of neurons must be appropriate. For the decorrelation model, the peak of the tuning curve determines the sampling distance that is required to achieve decorrelation.<sup>1</sup> If the response (i.e., the tuning curve multiplied with the image spectrum) is not flat, or if the sampling rate of the ganglion cell mosaic is too low, neighboring neurons will be correlated.

## 1.2. Ganglion cell correlations

Multineuron recording studies testing the independence of ganglion cell responses find that nearby ganglion cells of similar functional classes have significantly correlated firing patterns across species (Arnett, 1978; Arnett & Sparker, 1981; DeVries, 1999; Johnsen & Levine, 1983; Mastrojarde, 1989; Meister, 1996; Meister, Lagnado, & Baylor, 1995). There is also evidence that in development correlated firing is important for retinal neurons to innervate correctly: synchronized firing has been proposed as a mechanism that helps coordinate the proper development of neural wiring (see Wong, 2000). Therefore, if one wishes to argue that the primary goal of retinal coding is to produce a representation with uncorrelated responses, one must consider the evidence that the retina has not been fully successful.

## 1.3. Response equalization and vector length

The response equalization model does not make strong arguments regarding the particular shape of any individual tuning curves and it does not depend on the relative spacing of neurons. Our model argues that neurons have overall sensitivity set in such a way that different neurons have the same average response to a natural scene. For neurons tuned to different spatial frequency bands, those tuned to higher spatial frequencies must increase their overall sensitivity to counteract the  $1/f$  falloff in amplitude. In applying this argument to ganglion cells, we propose that the "integrated sensitivity" of neurons of different sizes across the retina is set in such a way that each neuron will respond approximately equally to natural scenes—despite the large differences in receptive field sizes and regions of pooling. In this paper, we focus on spatial properties but a full account would consider temporal tuning as well.

<sup>1</sup> If the tuning curve of the neuron increases with frequency out to some peak frequency  $P$  and falls off sufficiently fast past this point, then sampling at a frequency of  $2 \times P$  will produce uncorrelated responses in the presence of an image with a  $1/f^2$  power spectrum. Sampling at higher frequencies ( $> 2 \times P$ ) and therefore at smaller distances will result in correlated firing.

As an integrated measure of the sensitivity of a neuron, we used its “vector length” (Brady & Field, 1995; Field & Brady, 1997), i.e., the  $L_2$  norm of its sensitivity profile.<sup>2</sup> Fig. 1 shows two models for how neural sensitivity profiles, assumed to have constant shape, might scale with the peak spatial frequency of a neuron. On the left, peak sensitivity is independent of peak spatial frequency, so the  $L_2$  norm must increase with peak spatial frequency (“IVL” model). On the right, the  $L_2$  norm is constant with respect to peak spatial frequency (“CVL” model), so the peak sensitivity must decline with peak spatial frequency.

In a world with a  $1/f^2$  power spectrum, the vector length must increase with frequency in order to achieve response equalization (Brady & Field, 1995).

## 2. Study 1—calculation of vector length sensitivity of ganglion cells

In this first study, we investigate the hypothesis that at the level of the retina, ganglion cells already show evidence of response equalization. The study makes use of the published data of Croner and Kaplan (1995). That study is one of the few that provides a measure of the absolute sensitivity of ganglion cell sensitivity for a relatively wide range of receptive field sizes, as well as sufficient information for us to calculate the vector length. This first analysis is therefore a simple reanalysis of their data.

### 2.1. Methods

The Croner and Kaplan (1995) study measured responses of ganglion cells across the retina in anesthetized, paralyzed macaques when presented with gratings of different frequencies. In that study, the tuning functions were Fourier transformed assuming a center-surround phase spectrum, fit to a Difference of Gaussian (DoG) model and the median parameters of those center-surround neurons were published for various cell types and positions. Since linear transforms do not alter the vector length, we can use these data to calculate the vector length as a function of cell size. Absolute sensitivity data was collected across a number of animals and for both M- and P-cells.

In our study of the Croner and Kaplan (1995) data, we used their experimentally determined parameters for the DoG function describing the cells’ receptive fields (see Eq. (1)). We calculated the vector length ( $L_2$ -norm) of the DoG functions for P-cells in the study. P-cells are the dominant class found in the primate retina, and they have high spatial acuity compared to M-cells.

### 2.2. Results

We plotted our vector length sensitivity values as a function of the log (weighted) mean spatial frequency of each cell (i.e., the weighted mean value of the spatial frequency tuning curve of each cell<sup>3</sup>), shown in Fig. 2A. Parameters for the 84 total P-cells of different sizes represent the median value within bins corresponding roughly to cells of the same eccentricity on the retina. There are five such median values for the parameters that describe the receptive field function. The results of our analysis of the Croner and Kaplan data suggest that vector length is indeed increasing as a function of frequency.

What does this vector length sensitivity curve tell us about ganglion cell responses to natural scenes? With a power spectrum that falls as  $1/f^2$  (amplitude falls as  $1/f$ ), the response function should be approximately flat, indicating that the response from cells of different sizes to natural scenes will be approximately uniform. Fig. 2B shows the response of each neuron to a natural scene, computed by first multiplying each neuron’s tuning curve with a  $1/f$  amplitude spectrum and then taking the  $L_2$  norm of that product. This linear model suggests that P-cells perform a significant degree of response equalization.

Although we are ignoring the temporal aspects of the neural response and making the assumption that the system is linear, these results do imply that P cells have sensitivity that is well-matched to the power spectra of natural scenes. If these results hold for all P cells, then the prediction we make is that neurons of different sizes distributed across the retina will provide a roughly equal response.<sup>4</sup> Uniform responses across frequency implies that the cells are maximizing the use of the range of firing rates over which the cell responds, given the regular statistics of the environment.

We note that the data show no clear high-frequency cut-off but at this time, we cannot say whether the sensitivity continues to increase out to the highest spatial frequencies to which the neurons respond. It should also be noted that the vector length sensitivity makes a direct prediction regarding how the neurons will respond to noise. Without knowledge of how sensitivity was affected by mean luminance in the Croner and Kaplan (1995) study, we cannot say any more about the noise-reduction properties of primate ganglion cells in this study.

## 3. Study 2—decorrelation and sparseness in model ganglion cells

We stress that the response equalization hypothesis provides an account of the overall sensitivity of different

<sup>2</sup> The  $L_2$  norm of a vector  $\mathbf{R}$  is given by  $\sqrt{\sum_k R_k^2}$ , where  $R_k$  denotes the  $k$ th component of  $\mathbf{R}$ . Note that the vector length of  $\mathbf{R}$  will be the same in any orthonormal basis and it is therefore useful as a unitless, relative measure of sensitivity.

<sup>3</sup> Qualitatively similar results were obtained using the peak frequency value.

<sup>4</sup> M-cell data in the Croner and Kaplan study were insufficient for us to draw conclusions about patterns of vector length sensitivity at different spatial frequencies.

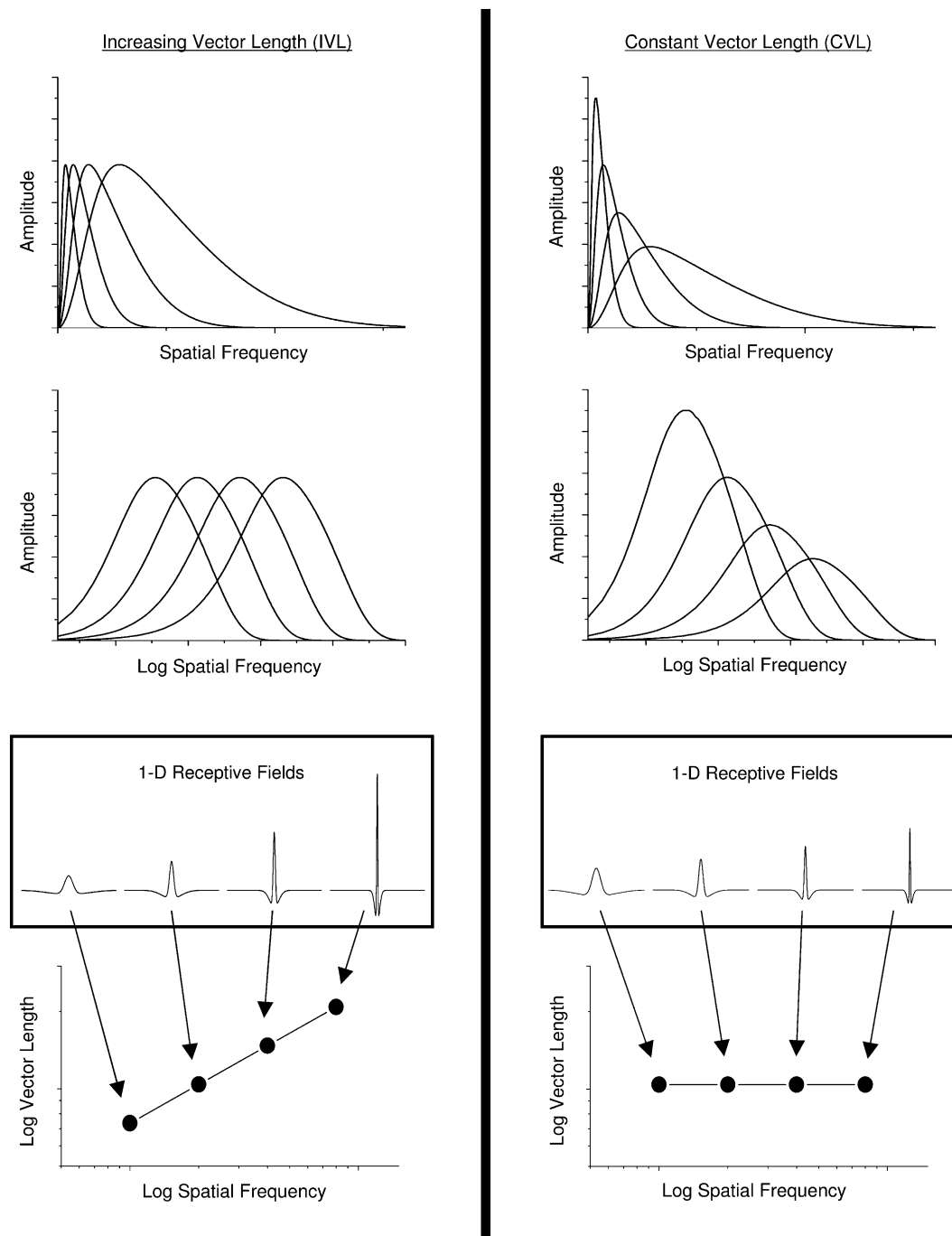


Fig. 1. A comparison of an increasing vector length model of contrast sensitivity, whose vector length increases with spatial frequency (left column), and a constant vector length model (right column), which shows equal vector length across frequencies. The tuning curves of four hypothetical retinal neurons are shown (from top to bottom) as functions of spatial frequency and log spatial frequency; as 1D receptive field profiles; and in terms of the log of their 1D vector length. In the increasing vector length model (IVL), the peak of each spatial frequency tuning curve is nearly the same, and thus neurons sensitive to high spatial frequencies integrate over proportionally more of the frequency spectrum. This model will show increasing vector length with increasing spatial frequency. The constant vector length model (CVL), on the other hand, predicts decreasing peaks in amplitude for increasing spatial frequencies and approximately equal vector length sensitivities across frequency. A full linear model of the contrast sensitivity function using the vector length metric is being developed (Field & Chandler, in preparation). For clarity in this figure, the vector lengths correspond to the 1D receptive field profiles shown. The 2D model discussed in the text is a straightforward extension of the 1D model shown in this figure. In both 1D and 2D, the vector length is given by its  $L_2$  norm. The 2D model used in this paper predicts that vector length increases proportional to spatial frequency (volume under the power spectrum increases in proportion to frequency); in the 1D model, vector length increases as the square root of spatial frequency.

neurons but has no implications regarding the spatial correlations between neighboring neurons. As was noted, there is significant evidence that there exist significant cor-

relations between neighboring neurons suggesting that the convolved spectra and/or the spacing is insufficient for producing decorrelated responses.

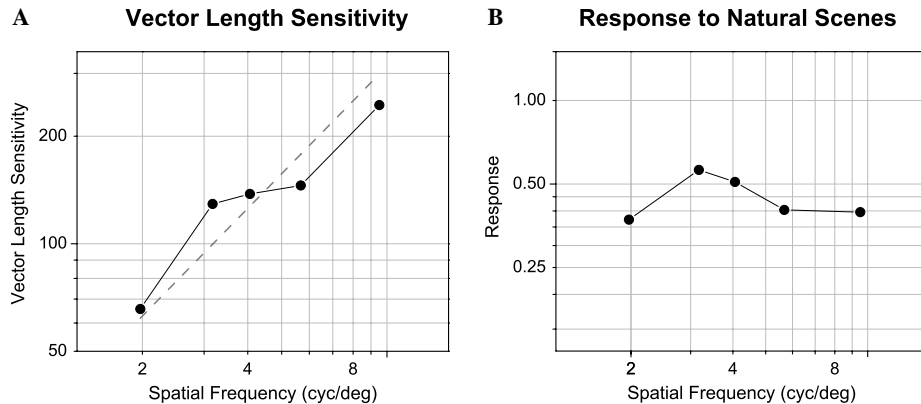


Fig. 2. (A) Plot of the vector length sensitivity of cells from data by [Croner and Kaplan \(1995\)](#). Each data point in the vector length plot (dots) represents a cell whose receptive field is modeled with the DoG parameters reported by [Croner and Kaplan \(1995\)](#). Vector length sensitivity (unitless) is monotonically increasing proportional to frequency. The  $x$ -coordinate of the vector length plot is in units of log mean spatial frequency (cyc/deg) for each cell (see text for definition). (B) Plot of the response magnitude for ganglion cells to a distribution with a spatial frequency power spectrum that falls as  $1/f^2$ . This  $1/f^2$  “input” represents a typical natural scene. Sensitivity is given by vector length (A). Responses show a generally flat shape across spatial frequency. Dotted line in (A) represents a slope of 1 on the log-log plot.

In this study, we wish to make a further point. We argue that both the decorrelation hypothesis and the response equalization hypothesis are dependent on the spectra of the neurons’ tuning curves and are not directly dependent on the phase spectra of the neurons. We argue that the center-surround organization, which depends on the phase spectra, is not directly addressed by either approach.

We focus on the question of what function is provided by localized center-surround receptive fields like those of retinal ganglion cells. To explain the oriented receptive fields of neurons in primary visual cortex, it has been argued that the visual system produces a sparse solution that reduces dependencies beyond the second-order correlations ([Field, 1987, 1994](#)). Neural networks that attempt to minimize these dependencies among the population of neurons have been found to produce localized, bandpass, oriented receptive fields much like those of simple cells found in V1 (e.g., [Bell & Sejnowski, 1997](#); [Olshausen & Field, 1996](#)). Therefore, if the goal of early coding were to produce an efficient or independent solution, we might expect to see a wavelet-like transform similar to V1 in the retina. A wavelet-like transform does not require more neurons than a center-surround system does to achieve a complete representation, so the argument cannot be that more neurons are needed.

Here, we investigate the relations among center-surround organization, decorrelation and sparseness in model retinal neurons. The simulation in the following section has two goals: first, it will be used to demonstrate that the decorrelation hypothesis is insufficient to predict center-surround receptive field design. Second, the simulation demonstrates that the ganglion cell produces a more sparse response than other solutions that decorrelate to the same extent.

### 3.1. Methods

For this study, images from van Hateren’s database ([van Hateren & van der Schaaf, 1998](#)) were randomly selected.

Images were then discarded if they did not conform to two criteria: They were required to be devoid of human-created forms and of significant blur. The restriction on blur is the more crucial one: if the camera moved when the shutter was open, the resulting images were blurry, which introduces uncertainties into the data. After our selection process, we arrived at a set of 137 stimuli that shows a range of scenes at different scales (images used are listed at <http://tinyurl.com/68mbb>). The mean power spectrum of the images was fit by the function  $y = 1/f^n$  where  $n = 2.6$ . This value of  $n$  reflects the fact that the images used are a biased data set within the [van Hateren and van der Schaaf \(1998\)](#) database. However, the relatively uniform ganglion cell response function described in the calculation above is qualitatively the same for  $n = 2.0$  and for  $n = 2.6$ , and the calculations below are not dependent on this fact. Most images in our dataset show grass or forest scenes, some have bodies of water and none has any large vistas. Calibrated images such as these are photometric maps of scenes wherein pixel values correspond linearly to luminance.

We used a difference-of-Gaussians (DoG) model of retinal ganglion cell receptive fields as the basis of our filter kernels (see [Fig. 3](#)). The DoG model is a simplified model that ignores many aspects of ganglion cell function. The radially symmetric DoG function  $R(x, y)$  is described by

$$R(x, y) = C_1 \frac{1}{\pi \sigma_1^2} e^{-(x^2+y^2)/\sigma_1^2} - C_2 \frac{1}{\pi \sigma_2^2} e^{-(x^2+y^2)/\sigma_2^2}, \quad (1)$$

where  $C_1$  and  $C_2$  are constants that determine the height of the center and surround Gaussians, respectively, and  $\sigma_1$  and  $\sigma_2$  are the variances of the center and surround, respectively ([Rodieck, 1965](#)). In our study,  $\sigma_1/\sigma_2 = 6.0$  and  $C = 20$  and filter kernels were created in a frame of  $64 \times 64$  pixels then centered and zero-padded to make them  $1024 \times 1024$  pixels (the size of the stimuli). Convolutions were performed with phase-aligned DoG filters and with DoG filters whose Fourier frequency components had been phase-randomized



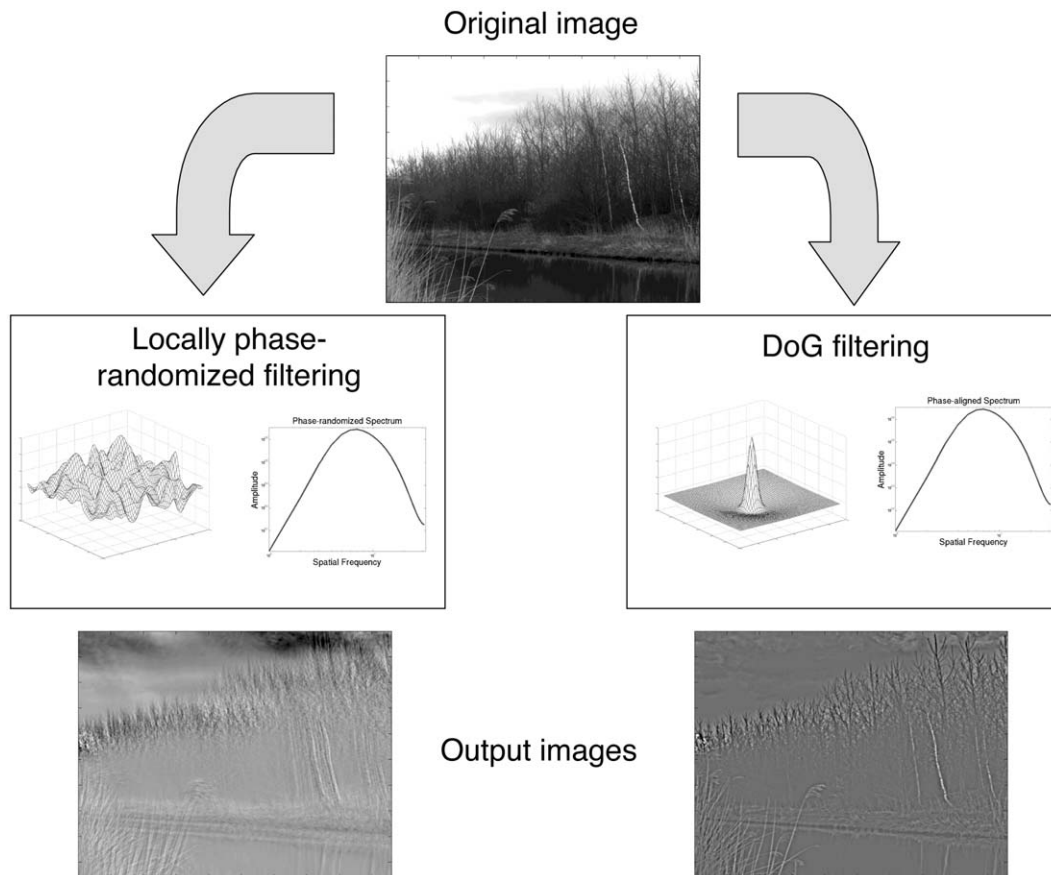


Fig. 3. Simulation of receptive field filtering of natural scene image. The original linear image (image number 6) was filtered with a phase-aligned difference of Gaussians (DoG) filter and with a locally phase-randomized filter (see text for definitions). The spectra of the two filters (shown on log-log coordinates) are identical. Note that in the output images some image structure is retained in the phase-randomized-filtered image as a result of the *localized* nature of the phase-randomized filters. We would not expect this to be the case if the filters were the same size as the image. (Filters are  $64 \times 64$  pixels before zero-padding and phase-randomization is done before zero-padding; images are  $1024 \times 1024$  pixels, though the images shown above are  $1024 \times 1536$ ). The phase-spectrum of the natural scene image was not manipulated before filtering.

(that is, they were given a norm-preserving, random rotation in the imaginary plane of phase space) before zero-padding. The term “phase-aligned” is used throughout to indicate that the frequency components of the filter are aligned at zero phase before zero-padding. The term “phase-randomized” is used throughout to indicate that Fourier frequency phases were randomized before zero-padding.

By necessity, the power spectra of the phase-aligned and the phase-randomized filters are identical.<sup>5</sup> Six phase-randomized filters were each convolved with the image set, and we took the mean of these trials as our phase-randomized power spectrum. The 64 pixels at the edges of the images were cropped before spectral analysis in order to remove edge effects (this was necessary for both types of filters because we did not use periodic boundary conditions).

### 3.2. Results

The first result should really be considered a mathematical necessity rather than an experimental finding. By randomizing the phase spectrum of the filter, we change the phase spectrum of the convolved image, but such randomization can have no effect on the resulting amplitude spectrum. The amplitude spectrum of the convolved image is simply the product of the amplitude spectrum of the image and the amplitude spectrum of the filter. The phase spectrum plays no role.

Since the autocorrelation function is the Fourier transform of the power spectrum, the phase spectrum also plays no role in determining the correlations. Phase-randomized filters achieve the same flattening in the 0.3–3 cycles/deg range and the same high-frequency noise attenuation as do center-surround filters. But the phase-randomized filters do not resemble ganglion cell receptive fields (see Fig. 3). For filters with a given power spectrum, each alignment of phases preserves the same amount of information in any convolution with an image.

We must therefore conclude that the center-surround structure does not follow from the constraint that the sys-

<sup>5</sup> The convolved power spectra are also identical when the image and the filter kernel are the same size, as we found in a separate trial. But in our experiment, because the images are larger than the filters, zero padding is necessary, which leads to differences in the convolved power spectra at low frequencies. However, the mean power spectrum for the set of images convolved with phase-randomized filters falls within one standard deviation of the mean power spectrum of images convolved with phase-aligned filters.

tem simply decorrelates. As with the response equalization hypothesis, there are a wide variety of solutions that achieve equivalent decorrelation, and the center-surround organization is just one example. We must therefore look to other constraints.

### 3.2.1. Sparseness

We measured the sparseness of our convolved images using kurtosis as our metric and found that the ratio of the sparseness of the center-surround-filtered images compared to that of images convolved with phase-randomized filters is on average  $3.5 \pm 0.40$  (mean  $\pm 95\%$  confidence limits). That is, the mean sparseness of center-surround filtered images is greater than the mean sparseness of phase-randomized-filtered images by a factor of  $3.5 \pm 0.40$  see Fig. 4. This value refers to the mean difference in *population* sparseness, defined as the sparseness across the population of neurons for a given static image. To gauge *lifetime* sparseness—that is, the sparseness of a neuron's response through its lifetime as opposed to the sparseness across the population—we compiled a total histogram for all images after filtering with phase-aligned and with phase-randomized filters. In this case, the center-surround filtered images had a sparseness that was  $1.4\times$  greater than that of images convolved with phase-randomized filters. As a control, the same convolutions with the two sets of filters were performed on Gaussian white noise and on white noise whose power spectrum was given by  $1/f^2$ , both of which gave a kurtosis of 0 for all convolutions.

### 3.2.2. Compressive non-linearities in the retina

In keeping with the proposal of Srinivasan et al. (1982), we convolved the same sets of filters with log-transformed images. The rationale for taking a log of the image before filtering is based on physiological studies of frogs by Norman and Werblin (1974), who showed that photoreceptor sensi-

tivity, when adaptation over time is taken into account, goes roughly as the log of intensity (see also Naka & Rushton, 1966; Baylor, Nunn, & Schnapf, 1987). Moreover, as Field (1987) pointed out, a log transform would recast intensity differences as ratios, a property that could be advantageous for the cell since intensity ratios express contrast.

We applied a log non-linearity to the images from the previous study then convolved each with the same sets of filters. Sparseness for the phase-aligned DoG filters was higher than the sparseness for the phase-randomized filters in this case by a factor of  $1.9 \pm 0.17$  see Fig. 4. We report the log case in order to show that sparseness is higher for the DoG filters than for the phase-randomized filters when a model of the cone non-linearity is included. Lifetime response sparseness for log-transformed images was found to be  $1.5\times$  greater for the center-surround filtered images than for those filtered with phase-randomized filters.

We note that the DoG filter and the phase-randomized filter are indistinguishable based solely on the mean response (the first statistical moment) because both filters were designed to have a mean of zero. Nor could they be distinguished based on variance (second moment): Because the filters have the same power spectrum, they will have the same variance. Differences in the skew (third moment) of the filtered images showed no clear pattern, whereas differences in kurtosis (fourth moment) did, as described above.

## 4. Discussion

This paper investigates several hypotheses regarding why the retina processes information as it does. One prejudice in the past has been to assume that because the retina is one of the earliest major processing units, it “is not expected to have knowledge beyond the simplest aspects of natural scenes” (Atick & Redlich, 1992). This has led many to assume that we have a relatively complete under-

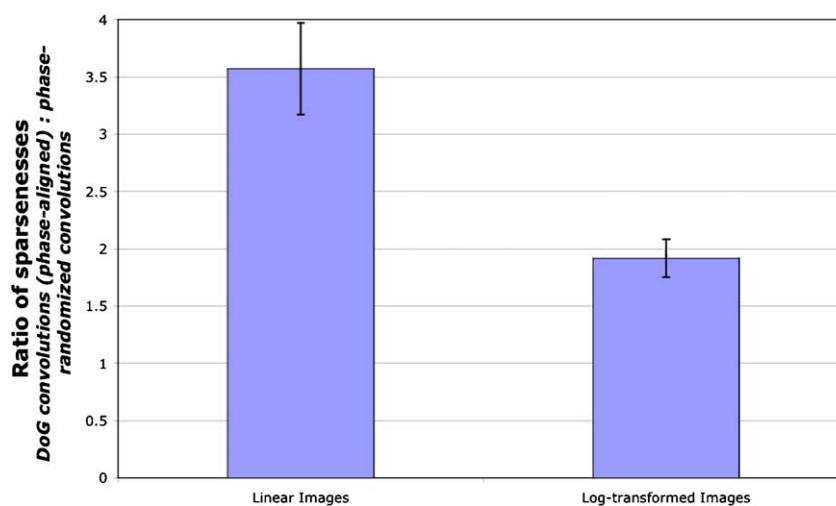


Fig. 4. Mean ratio across images of population sparseness for DoG convolutions compared to phase-randomized filter convolutions, for linear images (left bar) and for log-transformed images (right bar). For both types of images, DoG convolutions on average show greater sparseness per image. Error bars indicate 95% confidence limits. The mean sparseness ratio is the mean of the ratio of the sparseness for each pair of DoG and phase-randomized image convolutions, in both the linear and log-transformed cases.

standing of why the early visual system uses a center-surround receptive field. We argue that claims that the early visual system simply “whitens” or decorrelates the input are insufficient. To account for the center-surround organization, we believe that a set of at least three constraints is required.

Although evidence for full decorrelation among retinal neurons is lacking, we accept that decorrelation represents one of the constraints on the shape of the spatial frequency tuning curves and the relative spacing of cells. However, a decorrelation constraint does not account for the relative gain of different receptive field sizes, nor does it account for their center-surround organization. Moreover, the decorrelation hypothesis requires appropriate retinal sampling in order for its predictions to be valid.

Atick and Redlich (1992) make an important point about noise. Because the noise spectrum (e.g., photon noise) is thought to be flat and not declining in the same way as the signal (Pelli, 1981), decorrelation of a  $1/f$  spectrum in the high-frequency regime would serve to amplify unwanted noise. This argument is consistent with findings that ganglion cells lose their inhibitory surrounds at low luminance and become low-pass filters. As Atick and Redlich (1992) point out, low-pass filtering increases the signal-to-noise ratio because signal power becomes small at high frequency whereas noise power is constant across frequency. We accept this proposition but also wish to extend the argument. We believe that a full understanding of the underlying noise and vector length sensitivity can account for spatial sensitivity at threshold.

Using the vector length sensitivity measure, we find that sensitivity increases through at least 10 cycles/deg for P-cells in macaques. This result (as shown in Fig. 2) suggests that neurons with different sizes of receptive fields, from the fovea to the periphery, will respond about equally to a natural scene and maximize the use of the dynamic range available. In a previous study on psychophysical contrast matching it was proposed that this vector length sensitivity increases out to as much as 20 cycles/deg in humans (Brady & Field, 1995).

The vector-length approach to sensitivity may seem to conflict with the standard contrast sensitivity function (CSF) which implies that sensitivity peaks around 4 cycles/deg. However, we argue that there is no conflict. A full account of this argument is provided by Field and Chandler (in preparation). However, a brief comment should be made here.

#### 4.1. Contrast sensitivity and vector length

The contrast sensitivity function measures the psychophysical threshold at which humans (or any species) are able to detect contrast at a given spatial frequency. The CSF is fundamentally a signal-to-noise measure. If we presume that the noise that limits visual sensitivity is flat, then the vector length sensitivity is a direct measure of the noise magnitude in the system (Field & Brady, 1997). Further-

more, if we assume that the peak response to gratings is flat out to a range of 20 cycles/deg, and that the linear bandwidth increases with frequency (as shown in the left column of Fig. 1), then each neuron will have a constant response magnitude for its optimal grating but it will have a response to noise that increases with increasing frequency. The result is that the system will show a signal to noise level (i.e., the CSF) that decreases with increasing frequency, even though the vector length is increasing with increasing frequency. We propose that the 4 cycles/deg peak of the psychophysical contrast sensitivity function is the point at which the signal-to-noise ratio is maximized. The positive slope observed at low spatial frequencies could correspond to a regime that is coded by the lowest spatial frequency channel used by the visual system.

The results we show in Fig. 2 for P-cells are consistent with this general model. This simple linear model predicts the following:

1. Overall response sensitivity of neurons increases with increasing frequency (out to some limit—in the range of 20 cycles/deg in humans).
2. Response equalization. The response to natural scenes ( $1/f^2$  power spectrum) is roughly flat.
3. The contrast sensitivity function will fall at frequencies above and below the peak of the lowest channel.

As noted by Field and Brady (1997), this model also provides an account of why white noise appears to be dominated by high frequencies rather than structure at 4 cycles/deg. Unlike natural scenes or gratings, the model predicts that the response to noise peaks at the highest frequency channel (the neurons with the greatest vector length). If the CSF represented an accurate account of suprathreshold sensitivity, then one would expect that white noise would appear dominated by structure at 4 cycles/deg. A full account of how neural sensitivity relates to the CSF would also need to incorporate the role of the optics and the role of early non-linearities (Field and Chandler, in preparation). However, we wish to emphasize here that the CSF is not incompatible with this simple linear model where integrated sensitivity (vector length) is increasing.

#### 4.2. Further constraints

Our results suggest that both the decorrelation constraint and the response equalization constraint remain insufficient to predict the center-surround receptive field organization of ganglion cells. In response to natural scenes, we find that the center-surround organization of DoG filters produces a sparse response compared to phase-randomized filters. Since these two classes of filter achieve equal degrees of decorrelation and response equalization, we must consider that an additional constraint is required to account for the center-surround shape. Our simulation suggests that sparseness may be a factor. However, two points must be noted. First, if the only goal was



to represent the input with maximal sparseness, then the center-surround solution is not the optimal solution. Second, the sparseness may be partly a function of the highly localized nature of the center-surround profiles. As shown by Olshausen and Field (1996), a neural network that optimizes for sparseness and losslessness will settle on a set of oriented, bandpass filters similar to cortical simple cell receptive fields. If the only additional constraint were sparseness, we would expect to see an oriented wavelet code, which is not the case in the primate retina.

Recent physiological recordings in the primate retina (Berry, Warland, & Meister, 1997) and LGN (Reinagel & Reid, 2000) suggest that random flickering and white noise stimuli can both produce sparse responses in these cells' outputs. This implies some degree of non-linearity. We would expect sparseness to be lower for white noise than for natural scenes. But to our knowledge, no one has directly compared the sparseness of responses to natural scenes with the sparseness of responses to white noise.

Our findings apply to cells with receptive fields that are well-described by a DoG model (center-surround). Studies of ganglion cell sensitivity that use spike-triggered averages produced for white noise movies (white noise analysis) show that there could be additional microstructure in center-surround receptive fields that is not described by the DoG model (Brown, He, & Masland, 2000).

The fact that significant non-linearities exist in retinal processing (see Benardete & Kaplan, 1997a; Benardete & Kaplan, 1997b; Kaplan & Benardete, 2001; Shapley & Victor, 1979, 1981; Victor, 1987) could imply that these early non-linearities are good building blocks for the types of non-linearities found in the cortical code (end-stopping, cross-orientation inhibition, etc.). Retinal non-linearities could produce a code that is useful for the kinds of calculations performed in cortex. We have constrained ourselves to a linear model of ganglion cell spatial properties. However, we assume that non-linearities in the early visual system, including temporal and adaptive properties (Hosoya, Baccus, & Meister, 2005), may play an important role in ganglion cell tuning.

The answer to the question of why center-surround organization is highly conserved for receptive fields across species likely requires a broader theory that incorporates noise reduction, non-linearities, adaptation, and other temporal properties. Dong, among others, (Dong & Atick, 1995a, 1995b; Dong, 2001) has emphasized spatiotemporal decorrelation as an important goal of ganglion cells and of LGN. Linsker (1989) proposed an unsupervised learning algorithm that was optimized with respect to mutual information (equivalently, decorrelation)—this system could produce topographic maps, lateral interactions and Hebb-like modification, though not center-surround receptive fields.

### 4.3. Localization

It is possible that a center-surround arrangement requires a minimum of dendritic wiring given its task. Such

arguments have been considered in the context of cortex (e.g., Mitchison, 1991), though such arguments do not address receptive field organization specifically. In our simulation, one effect of phase-randomization was to increase the radial spread of the receptive fields in space, which results in a less sparse response (see Fig. 3).

Vincent and Baddeley (2003), using a set of simulations, argue that the center-surround operator serves to optimize synaptic efficiency. Presumably, minimizing the dendritic spread will serve to both minimize the total wiring needed in the retina and the number of synapses required to represent the input. From this line of argument, center-surround organization acts to optimize the localization of the receptive field for a given frequency tuning curve. That is, given some constraints on the tuning curve—for example, that it is required to be unimodal—center-surround organization achieves optimal localization (see Table 1).

One possible test of the minimum wiring hypothesis in ganglion cells would involve a neural network architecture that is designed to search for receptive fields that decorrelate and/or sparsify with a minimum of connectivity. We believe that both sparseness/independence and size/efficiency constraints are inter-related. Both serve useful goals. Although we believe that the size constraint may be the more important factor, the additional sparseness/independence should not be ignored. Given that current models are insufficient, the many factors that we believe could influence the goals of retinal processing are summarized in Table 1.

### 4.4. Conclusion

The fact that retinal structure and organization are remarkably well-conserved across mammalian species could imply that this organization is a very efficient first step in coding the natural world given the constraints of retinal neurophysiology. Moreover, to the extent that retinal processing is an optimally efficient first step for coding natural scenes, artificial visual systems may benefit from adopting a retina-like strategy as a first step as well. Such a strategy may prove useful as an initial stage in the extraction of features from many classes of natural images. Furthermore, the same types of constraints that contribute to center-surround organization in retinal ganglion cells (sparseness, response equalization, localization, minimal wiring, and other factors) may well explain the center-surround receptive fields in other sensory modalities (such as the tactile system) and the lateral inhibition found in the auditory system.

We conclude that a minimum of three constraints must be considered to account for the known linear properties—decorrelation, response equalization, and size/sparseness. Although we accept that decorrelation plays a role, the evidence does not support the hypothesis that the retina successfully decorrelates. Our work with P-cells from the primate study of Croner and Kaplan (1995) suggests that

Table 1  
Theories of efficient retinal coding

Goal	Explanation	Can this account for center-surround RFs?
Compression (lossy)	<ul style="list-style-type: none"><li>• To meet the reduced capacity of the optic nerve, attenuation of high spatial frequencies in periphery provides an efficient means of reducing bandwidth between photoreceptors and ganglion cells. The highest frequencies have the lowest signal-to-noise ratio given a <math>1/f</math> environment and flat noise</li><li>• Noise is reduced in regime where noise and signal power approach equivalence</li></ul>	<ul style="list-style-type: none"><li>• A wide variety of transforms can achieve high-frequency attenuation</li><li>• Insufficient constraint to account for center-surround receptive fields</li></ul>
Decorrelation	<ul style="list-style-type: none"><li>• Cells with tuning functions that rise with frequency can reduce pairwise correlations when the cells are spaced appropriately. The known tuning functions should therefore help reduce this form of statistical redundancy</li></ul>	<ul style="list-style-type: none"><li>• Many transforms can achieve decorrelation. Even if the tuning curve was optimal for decorrelation, many possible transforms can have the same tuning curves</li><li>• Visual system maintains significant correlations between neighboring neurons, and these correlations may be important in development</li><li>• Insufficient constraint to account for center-surround RFs</li></ul>
Response equalization	<ul style="list-style-type: none"><li>• Allows neurons of all sizes to produce the same average response to natural scenes with <math>1/f</math> spectra. Predicts response magnitude of macaque P cells</li></ul>	<ul style="list-style-type: none"><li>• A variety of transforms can sphere <math>1/f</math>-distributed data</li><li>• Insufficient on its own to account for center-surround receptive fields or shape of tuning curves</li></ul>
Sparseness	<ul style="list-style-type: none"><li>• Center-surround RFs are more sparse than phase-randomized RFs</li></ul>	<ul style="list-style-type: none"><li>• Wavelet-like/oriented receptive fields can provide greater sparseness than center-surround receptive fields</li></ul>
Minimal size/wiring	<ul style="list-style-type: none"><li>• Could achieve most sparse solution and reduce total dendritic field needed</li><li>• Restricts phases towards an aligned center-surround organization</li></ul>	<ul style="list-style-type: none"><li>• A minimal wiring constraint on its own cannot produce an extended receptive field</li></ul>
Decorrelation + response equalization + minimal size	<ul style="list-style-type: none"><li>• The combination of all three constraints may be sufficient to describe the basic sensitivity and center-surround organization</li></ul>	<ul style="list-style-type: none"><li>• Further consideration is needed to account for known non-linearities and temporal properties</li></ul>

It should be emphasized that this analysis largely ignores the importance of non-linearities in retinal processing.

the sensitivity across neurons of different sizes serves to produce equalized responses in the presence of natural scenes. A full account certainly must consider the temporal aspects of tuning, and other classes of retinal ganglion cell. However, our results argue that constraints on dendritic wiring and sparseness must also be considered. An account of the retina's linear functional goals would consider all of these factors and perhaps others. We emphasize that our approach has not considered the range of non-linearities found across different classes of retinal neurons. How many more constraints will be required to provide a full account of retinal processing remains to be seen.

### Acknowledgments

This work was supported by the following grants and contracts: D.J.G. was supported by NIH EY015393 (Kirs-

chstein-NRSA) and NSF DGE-9870631 (IGERT Program in Nonlinear Systems fellowship); D.J.F. and D.M.C. were supported by NGA Contract HM 1582-05-C-0007 to D.J.F.

### Appendix A. Theory

It can be helpful to think of the two hypotheses of retinal coding—decorrelation and response equalization—in terms of vector spaces. First consider the case of two-dimensional Gaussian data with a strong correlation between two orthogonal vectors (Fig. 5A), for example, two pixels. One method for generating a decorrelated representation is to perform principal components analysis (PCA) on the input data. This method produces an orthogonal vector space whose axes are aligned with the directions along which Gaussian data have the highest variance. The vectors

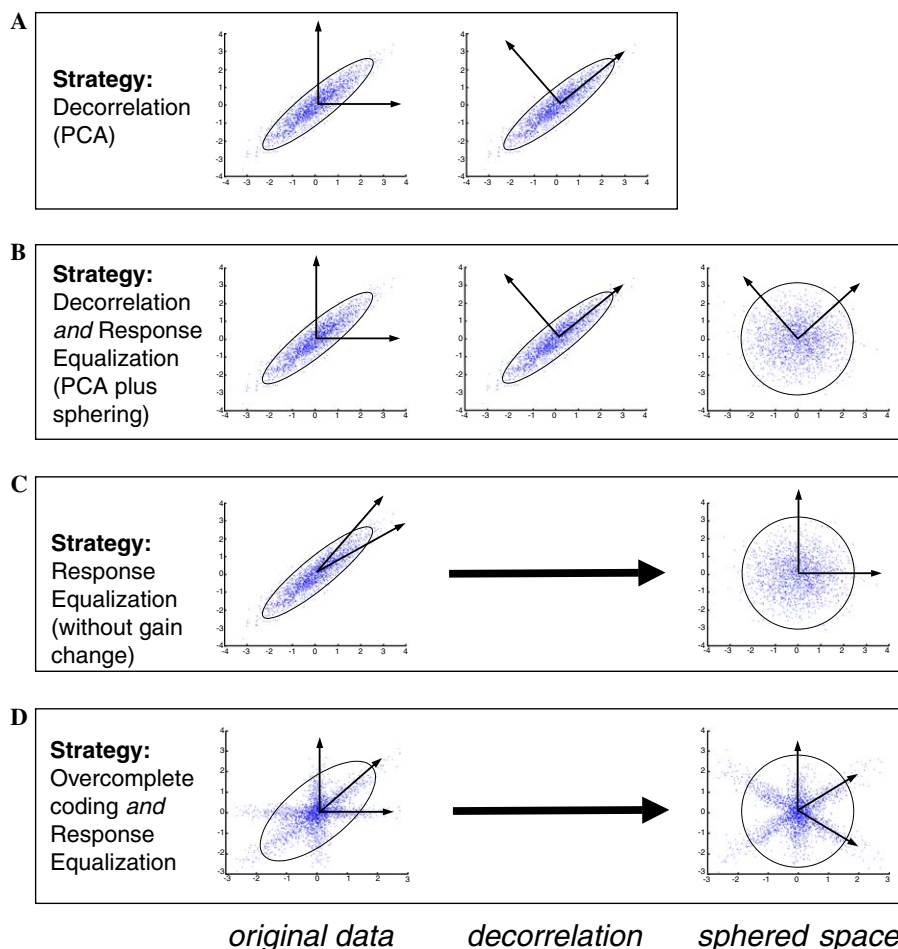


Fig. 5. Vector space representations of efficient coding strategies. (A) Shows a two-dimensional PCA transform on Gaussian data. The axes are rotated so that they are aligned with the principal components of the data. In (B), the principal component vectors are normalized or “sphered” (B, right panel) such that the variance along each of the basis vectors is normalized. In (C), we show an example of a response equalization coding strategy without a change in vector length. The two vectors shown in the untransformed space (C, left panel) are not orthogonal. The correct choice of vectors results in response equalization. The outcomes of the transforms in both (B and C) are part of a family of rotations in the sphered space. The data in (D) lie in a two-dimensional space but in order to discover the six-pointed star-shape of the data, three non-orthogonal basis vectors are required. The superimposed ellipse in (D) has its major and minor axes aligned with the two principal axes of these data. An over-complete coding strategy that includes response equalization aligns the representation's basis vectors with the causes of the data (D, right panel) but these vectors are not uncorrelated (although they are sparse), even in the sphered space.

generated by PCA will be uncorrelated but their response variance will not be equal.

The principal component vectors can be normalized or “sphered” as shown in Fig. 5B such that the variance along each of the basis vectors is normalized. This normalization—also called response equalization—allows all vectors (or neurons) to respond with the same average magnitude to the family of inputs. This combination of PCA and response equalization is sometimes referred to as “sphering” or “whitening.” However, both terms can be misleading. Sphering is not part of PCA, but for the example shown in Fig. 5B, sphering produces a representation whose variance is normalized with respect to the basis vectors (that is, response equalized), thus creating a univariate Gaussian distribution. It should be noted that this process creates a sphere only when one is given Gaussian data. If the data are not Gaussian (as shown in Fig. 5D), the sphering will result in both decorrelation and response equalization, but there will remain higher-order statistical dependencies.

Fig. 5C demonstrates another way to achieve sphering. By choosing the right set of non-orthogonal axes, one can achieve both decorrelation and response equalization for these data. In the sphered space shown on the right of Figs. 5B and C, the two transforms are simply rotations of one another. In Fig. 5C, the gain of the two neurons is the same but the response is effectively sphered. It is therefore theoretically possible to sphere data without a gain change.

However, ganglion cells with different size receptive fields will not see the same stimulus strength. Because of the  $1/f^2$  power spectrum, the neurons tuned to higher spatial frequencies (smaller receptive fields) will see less signal strength. As we will argue, in order to achieve equalization in response to natural scenes the neurons with smaller receptive fields must increase their gain. The response equalization hypothesis suggests that the relative gain of neurons tuned to different frequencies is designed to equalize the response of neurons of different sizes. The hypothesis makes no assumptions about the amount of decorrelation achieved by retinal processing. Consider a case where the causes of the data are non-orthogonal and let us assume that the number of causes is over-complete (i.e., there are more causes of the data than there are dimensions in the representation), as in Fig. 5D. There exists no linear transform of these data that will result in independent responses. We might choose to align our vectors with the causes of the data as shown in Fig. 5D. If the causes are not orthogonal, the vector outputs will be correlated. But regardless of the correlations, it may be desirable to perform response equalization: for these data, the different causes have unequal variance so vectors of the same length aligned with these causes will therefore have unequal outputs. However, if the vector lengths are adjusted to counter this difference in variance, response equalization is possible (even though the correlations will remain) as shown in the figure.

We are therefore left with the possibility of achieving decorrelation with or without response equalization—and of achieving response equalization with or without decorrelation.

## Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.visres.2006.03.008](https://doi.org/10.1016/j.visres.2006.03.008).

## References

- Atick, J. J., & Redlich, A. N. (1992). What does the retina know about natural scenes? *Neural Computation*, 4, 196–210.
- Arnett, D. W. (1978). Statistical dependence between neighboring retinal ganglion cells in goldfish. *Experimental Brain Research*, 32, 49–53.
- Arnett, D., & Sparker, T. E. (1981). Cross-correlation analysis of maintained discharge of rabbit retinal ganglion cells. *Journal of Physiology*, 317, 29–47.
- Balboa, R. M., & Grzywacz, N. M. (2000). The role of early retinal lateral inhibition: more than maximizing luminance information. *Visual Neuroscience*, 17, 77–89.
- Baylor, D. A., Nunn, B. J., & Schnapf, J. L. (1987). Spectral sensitivity of the cones of the monkey *Macaca fascicularis*. *Journal of Physiology*, 390, 145.
- Bell, A. J., & Sejnowski, T. J. (1997). The independent components of natural scenes are edge filters. *Vision Research*, 37, 3327–3338.
- Benardete, E. A., & Kaplan, E. (1997a). The receptive field of the primate P retinal ganglion cell, I: linear dynamics. *Visual Neuroscience*, 14, 169–186.
- Benardete, E. A., & Kaplan, E. (1997b). The receptive field of the primate P retinal ganglion cell, II: nonlinear dynamics. *Visual Neuroscience*, 14, 187–205.
- Berry, M. J., Warland, D. K., & Meister, M. (1997). The structure and precision of retinal spike trains. *Proceedings of the National Academy of Sciences of the United States of America*, 94, 5411–5416.
- Brady, N., & Field, D. J. (1995). What's constant in contrast constancy: the facts of scaling on the perceived contrast of bandpass patterns. *Vision Research*, 35, 739–756.
- Brady, N., & Field, D. J. (2000). Local contrast in natural images: normalisation and coding efficiency. *Perception*, 29, 1041–1055.
- Brown, S. P., He, S., & Masland, R. H. (2000). Receptive field microstructure and dendritic geometry of retinal ganglion cells. *Neuron*, 27, 371–383.
- Burton, G. J., & Moorhead, I. R. (1987). Color and spatial structure in natural scenes. *Applied Optics*, 26, 157–170.
- Croner, L. J., & Kaplan, E. (1995). Receptive fields of P and M ganglion cells across the primate retina. *Vision Research*, 15, 7–24.
- DeVries, S. H. (1999). Correlated firing in rabbit retinal ganglion cells. *Journal of Neurophysiology*, 81, 901–920.
- Dong, D. W., & Atick, J. J. (1995a). Statistics of natural time-varying images. *Network Computations Neural Systems*, 6, 345–358.
- Dong, D. W., & Atick, J. J. (1995b). Temporal decorrelation: a theory of lagged and nonlagged responses in the lateral geniculate nucleus. *Network Computations Neural Systems*, 6, 159–178.
- Dong, D. W. (2001). Spatiotemporal inseparability of natural images and visual sensitivities. In J. M. Zanker & J. Zeil (Eds.), *Motion vision: Computation, neural, and ecological constraints*. Berlin: Springer Verlag.
- Field, D. J. (1987). Relations between the statistics of natural images and the response profiles of cortical cells. *Journal of Optical Society America*, 4, 2379–2394.
- Field, D. (1994). What is the goal of sensory coding? *Neural Computations*, 6, 559–601.

- Field, D. J., & Brady, N. (1997). Wavelets, blur and the sources of variability in the amplitude spectra of natural scenes. *Vision Research*, 37, 3367–3383.
- Field, D.J., Chandler, D.M. 2006. Where is the peak of visual sensitivity? In preparation.
- Finlay, B. L., de Lima Silveira, L. C., & Reichenbach, A. (2004). Comparative aspects of visual system development. To appear. In J. Kremers (Ed.), *The structure, function and evolution of the primate visual system*. New York: John Wiley.
- Hosoya, T., Baccus, S. A., & Meister, M. (2005). Dynamic predictive coding in the retina. *Nature*, 436, 71–77.
- Johnsen, J. A., & Levine, M. W. (1983). Correlation of activity in neighbouring goldfish ganglion cells: relationship between latency and lag. *Journal of Physiology*, 345, 439–449.
- Kaplan, E., & Benardete, E. (2001). The dynamics of primate retinal ganglion cells. *Progress in Brain Research*, 134, 17–34.
- Land, M. (1985). The eye: optics. In G. A. Kerkut & L. I. Gilbert (Eds.), *Comprehensive insect physiology, biochemistry and pharmacology*. London: Pergamon.
- Linsker, R. (1989). How to generate ordered maps by maximizing the mutual information between input and output. *Neural Computations*, 1, 402–411.
- Mastronarde, D. N. (1989). Correlated firing of retinal ganglion cells. *Trends in Neurosciences*, 12, 75–80.
- Meister, M. (1996). Multineuronal codes in retinal signaling. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 609–614.
- Meister, M., Lagnado, L., & Baylor, D. A. (1995). Concerted signaling by retinal ganglion cells. *Science*, 270, 1207–1210.
- Mitchison, G. (1991). Neuronal branching patterns and the economy of cortical wiring. *Proceedings of the Royal Society of London Series B. Biological Sciences*, 245, 151–158.
- Naka, K. I., & Rushton, W. A. (1966). S-potentials from colour units in the retina of fish (Cyprinidae). *Journal of Physiology*, 185, 536–555.
- Norman, R. A., & Werblin, F. S. (1974). Control of retinal sensitivity. I. Light and dark adaptation of vertebrate rods and cones. *Journal of General Physiology*, 63, 37–61.
- Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381, 607–609.
- Pelli, D.G. 1981. Effects of visual noise. Ph. D. thesis. Cambridge University, Cambridge, England.
- Ratliff, F. (1965). *Mach bands: Quantitative studies on neural networks in the retina*. San Francisco: Holden-Day.
- Reinagel, P., & Reid, R. C. (2000). Temporal coding of visual information in the thalamus. *Journal of Neurosciences*, 20, 5392–5400.
- Rodieck, R. W. (1965). Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research*, 5, 583–601.
- Shapley, R., & Victor, J. D. (1979). Nonlinear spatial summation and the contrast gain properties of cat retinal ganglion cells. *Journal of Physiology*, 290, 141–161.
- Shapley, R., & Victor, J. D. (1981). How the contrast gain control modifies the frequency responses of cat retinal ganglion cells. *Journal of Physiology*, 318, 161–179.
- Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. *Proceedings of Royal Society of London Series B. Biological Sciences*, 216, 427–459.
- van Hateren, J. H., & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of Royal Society of London Series B. Biological Sciences*, 265, 359–366.
- Victor, J. D. (1987). The dynamics of cat retinal X cell centre. *Journal of Physiology*, 386, 219–246.
- Vincent, B. T., & Baddeley, R. J. (2003). Synaptic energy efficiency in retinal processing. *Vision Research*, 43, 1283–1290.
- Wong, R. O. L. (2000). Retinal waves and visual system development. *Annual Review of Neuroscience*, 22, 29–47.